Détection de caractéristiques locales invariantes aux changements affines

Krystian Mikolajczyk

Résumé de la thèse

La reconnaissance des formes dans les images est un problème central de l'intelligence artificielle. La quantité abondante de données numériques nécessite des outils automatiques et performants pour la reconnaissance et la recherche de l'information pertinente. L'objectif principal de la reconnaissance des formes est d'extraire le contenu sémantique des images. Une des approches dominantes est basée sur les caractéristiques locales. La méthode utilise la description locale calculée au voisinage de points d'intérêt. La détection de points d'intérêt est une première étape dans le processus de la mise en correspondance et de la reconnaissance. L'approche par apparences locales a permis d'améliorer et d'accélérer considérablement la recherche d'images dans des bases de données. Dans cette thèse, nous abordons le problème d'extraction de points caractéristiques d'images qui sont ensuite utilisés pour apparier les images. La deuxième partie de la recherche s'inscrit dans le domaine de la reconnaissance de classes d'objets.

Appariement d'images. Nous proposons une nouvelle approche pour la détection de points caractéristiques d'une image. Cette approche est invariante aux transformations géométriques et photométriques, qui apparaissent fréquemment entre les images prises dans des conditions différentes. Nous nous concentrons sur le problème d'invariance aux transformations affines. Cette transformation est particulièrement importante parce qu'elle permet de s'affranchir des problèmes de changements perspectives. Les approches précédentes apportent des solutions partielles, car certains paramètres de points d'intérêt ne sont pas estimés de façon invariante aux changements affines. Nous avons proposé une solution générique à ces problèmes.

La première méthode proposée dans la thèse permet d'extraire des points d'intérêt de façon invariante aux changements d'échelle. Nous avons proposé d'utiliser la sélection automatique d'échelle pour déterminer la taille du voisinage d'un point d'intérêt. Les points caractéristiques sont indiqués par des maxima locaux dans l'espace d'échelle construit à partir de dérivées normalisées. Pour un point donné on calcule des réponses de dérivées pour plusieurs facteurs d'échelle. Ainsi, on construit la représentation en échelle pour le voisinage de ce point. Dans l'ensemble des réponses, on cherche une échelle caractéristique relativement indépendante de la résolution de l'image. Cette échelle caractéristique est indiquée par un maximum local des réponses. Nous avons mené une étude théorique et une évaluation expérimentale de la technique de sélection automatique d'échelle. Cette évaluation montre que la mesure du Laplacien permet d'obtenir les meilleurs résultats pour la sélection de l'échelle caractéristique.

La deuxième approche est invariante aux changements affines y compris aux changements d'échelle importants. Nous proposons une solution pour la détection de points d'intérêt dans l'espace affine construit avec des filtres gaussiens affines. La méthode estime de façon itérative la déformation affine du voisinage d'un point. La localisation de points est effectuée par le détecteur de Harris basée sur la

matrice du deuxième moment. L'échelle caractéristique est indiquée par les maxima locaux de la mesure du Laplacien. Etant donné le point et l'échelle initiale, nous utilisons les propriétés de la matrice de deuxième moment pour déterminer la région affine centrée dans le point d'intérêt. Une procédure itérative est appliquée à chaque point initial fourni par le détecteur Harris multi-échelle. Cette procédure modifie simultanément la position, l'échelle ainsi que le voisinage affine du point. La méthode permet de converger vers un point stable qui est réellement invariant aux changements affines, même si le point initiale est relativement loin de la solution finale.

Les images sont caractérisées par des ensembles de descripteurs calculés en des points caractéristiques détectés automatiquement. Pour caractériser un point d'intérêt nous utilisons des dérivées de niveaux de gris calculées jusqu'à l'ordre 4. Afin d'obtenir des dérivées indépendantes de la rotation existante entre deux images, la direction de calcul des dérivées est rapportée à la direction dominante dans le voisinage d'un point. Ces descripteurs permettent ensuite d'indexer des images en étant invariant aux rotations, translations et changements affines même importants. Une mesure de ressemblance permet d'établir des correspondances entre les points des images prises de points de vue différents. Ces correspondances sont ensuite utilisées pour calculer la géométrie qui lie les images. Dans le contexte de la recherche d'images les descripteurs sont utilisés pour retrouver des points similaires dans la base et par conséquent des images similaires aux images requêtes. Les résultats expérimentaux pour la mise en correspondance et la recherche d'images montrent que notre approche est très robuste et efficace même dans les cas de changements importants. Nous avons constaté que la méthode invariante à l'échelle donne des résultats fiables jusqu'à un facteur d'échelle de 4,4 et la méthode affine jusqu'a un changement de point de vue de 70 degrés. Plusieurs études comparatives effectuées dans cette thèse montrent l'avantage de notre méthode par rapport aux approches existantes présentées récemment dans la littérature.

Reconnaissance de classes d'objets. La deuxième partie de la thèse est consacrée à la reconnaissance de classes d'objets. La détection de visages est un excellent exemple de la classification. L'objectif de la détection est de déterminer la localisation, la taille et la pose du visage dans une image. Nous avons développé une méthode innovante de détection des visages dans la séquence vidéo. Le détecteur de visage intègre les avantages de la détection dans des images fixes et l'aspect temporel d'une vidéo. Ceci permet d'améliorer la détection et d'effectuer le suivi de visages dans des séquences vidéo. L'approche est basée sur deux modèles: un pour les visages et l'autre pour les objets différents du visage. Les attributs caractéristiques des visages sont décrits par les distributions de coefficients obtenus par la transformation en ondelettes. Une image requête est comparée avec le modèle du visage et ensuite avec le modèle du non-visage. Une mesure de similarité permet de classifier l'image requête selon le résultat de cette comparaison. Les contributions principales de ce travail sont: l'accumulation des probabilités de détection dans une séquence pour obtenir une détection cohérente au cours du temps, la prédiction des paramètres de position, de l'échelle (taille) et de la pose du visage. Un des problèmes difficiles dans la détection de visages est le changement de pose du visage. Nous proposons d'utiliser deux détecteurs, un pour les visages vus de face et l'autre pour les profils, ainsi que l'information temporelle dans la séquence. La combinaison de ces deux détecteurs permet de s'affranchir de problèmes de changements de pose. Les résultats expérimentaux montrent une amélioration importante par rapport à une détection instantanée.

Mots Cles: Points d'intérêt, détection de points caractéristiques, invariance affine, description locale, mise en correspondance, recherche d'images, reconnaissance par apparence locale .